PSYCHOMETRIKA—VOL. 54, NO. 3, 501–513
SEPTEMBER 1989

ROBUST MULTIDIMENSIONAL SCALING

IAN SPENCE AND STEPHAN LEWANDOWSKY

UNIVERSITY OF TORONTO

A method for multidimensional scaling that is highly resistant to the effects of outliers is described. To illustrate the efficacy of the procedure, some Monte Carlo simulation results are presented. The method is shown to perform well when outliers are present, even in relatively large numbers, and also to perform comparably to other approaches when no outliers are present.

Key words: multidimensional scaling, robust estimation, outliers.

1. Introduction

In recent years, much attention has been paid to the topic of robust estimation. This is partly due to the fact that there is a considerable body of evidence to suggest that estimation techniques which presume Gaussian errors—for example, classical least squares—do not perform well when the data are contaminated by outliers. Proximity judgments, which form the basic data for most applications of multidimensional scaling procedures, may be particularly susceptible to the problem of outlying observations. In addition to the usual difficulties of detecting and eliminating transcriptional and data entry errors, there is often the problem of subject boredom while making large numbers of judgments, leading to erratic performance in some cases. Another not uncommon error may occur when a subject occasionally misuses a rating scale, perhaps inadvertently reversing the direction. Further, some judgments are intrinsically more difficult to make and it is not unreasonable to suppose that the error distributions associated with these have larger variances. Consequently, instead of dealing with a single error distribution—as is at least implicitly assumed by most procedures—we may be faced with a mixture of distributions, whether Gaussian or not. Therefore, Spence (1982), Null and Sarle (1982), and Heiser (1987) have argued the need for multidimensional scaling algorithms that are resistant to the effect of outliers.

As demonstrated below, even a single outlier may dramatically distort a multidimensional scaling solution when traditional metric scaling is employed. It is also shown that other varieties of multidimensional scaling may be adversely affected by outliers. However, a robust multidimensional scaling program, based on the algorithm described below, is shown to be little affected by outliers, even in relatively large numbers, and, furthermore, seems to perform adequately when no outliers are present.

2. The Fragility of the Traditional Method of Least Squares

The oldest method of multidimensional scaling is based on the work of Young and Householder (1938) and Torgerson (1958). It is also probably the most widely used method because, although nonmetric procedures may have largely supplanted classical

This research was supported by Grant A8351 from the Natural Sciences and Engineering Research Council of Canada to Ian Spence.

Requests for reprints should be sent to the first author at the Department of Psychology, University of Toronto, Toronto, Ontario, CANADA M5S 1A1.

0033-3123/89/0900-6031$00.75/0

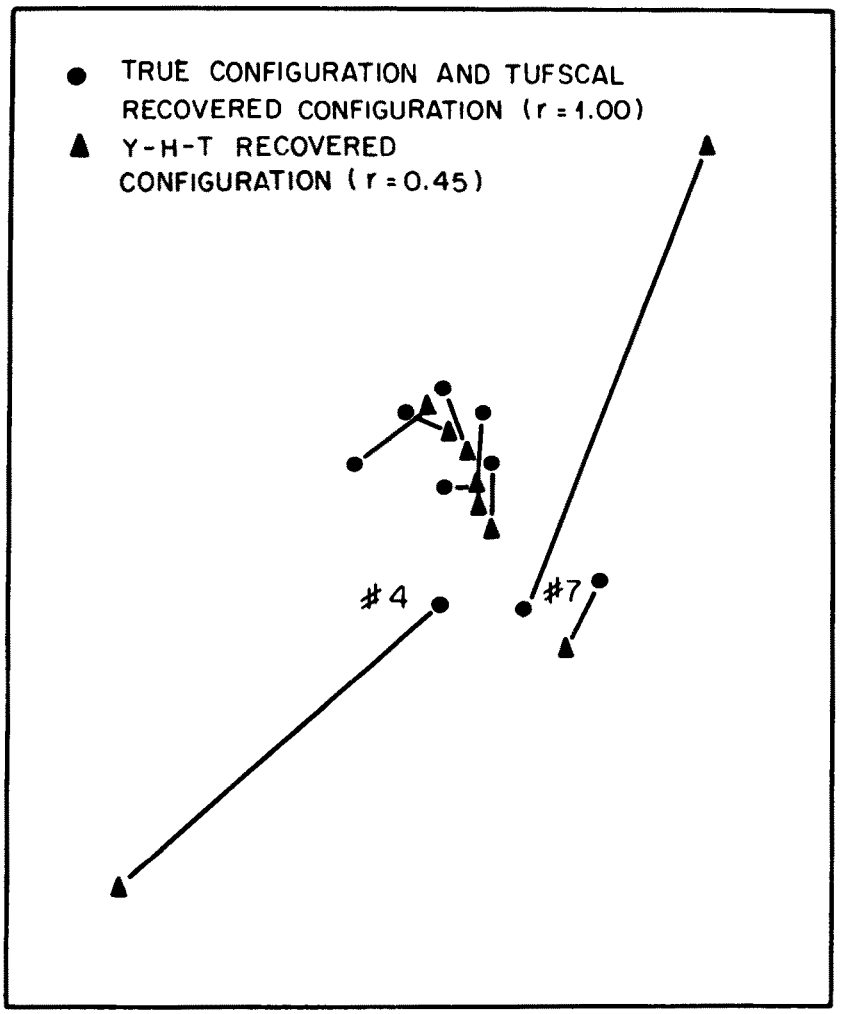

FIGURE 1
Recovery of perfect data with a single simulated data entry error (d_{47} multiplied by 10) using classical metric scaling.

metric scaling, almost all the currently popular programs employ some variant of the Young-Householder-Torgerson procedure as a starting configuration. Obviously, if the traditional approach is sensitive to aberrant data points, then this has implications for programs that use it as their first approximation to a solution.

A single outlier in otherwise perfect data may be simulated by multiplying a randomly chosen proximity by a constant to mimic the effect of incorrect column entry. A configuration consisting of nine points randomly selected in the unit plane was used to create a matrix of interpoint distances to be used as input to a metric multidimensional scaling routine. One of the distances, specifically d_{47}, was perturbed by multiplying it by 10, before scaling the matrix. These data, perfect in every respect bar one, were subjected to the Young-Householder-Torgerson procedure and the results are shown in Figure 1.

The effect of a single outlier is disastrous, and the result is precisely analogous to what can happen with almost perfectly linear data in ordinary least squares. As demonstrated by the "Andrews case" (see Mosteller & Tukey, 1977, p. 212), a single bad data point can exert considerable leverage on the fitted line, rendering least squares

estimates of the relevant parameters virtually useless. With bivariate data the aberration is easily detected in the scatterplot, but is rather more difficult to discover when there are several explanatory variables. Similarly, if outliers are present in multidimensional scaling, some work is necessary to discover the problem—at a minimum, the distribution of residuals must be carefully examined, but even then the outliers may not be very evident (Huber, 1981, p. 155).

Since the classical method is so fragile, it seems sensible to try to find more resistant techniques.

3. An Algorithm for Robust Parameter Estimation

In two way multidimensional scaling, we desire a representation such that $*d_{ij} - d_{ij} = e_{ij}$, where $*d_{ij}$ is the dissimilarity between objects i and j, and d_{ij} is the distance between points i and j:

$$d_{ij}^2 = \sum_{a=1}^{m} (x_{ia} - x_{ja})^2,$$

where a indexes the m dimensions of the euclidean representation and x_{ia} is the coordinate of the i-th point on the a-th dimension. The e_{ij} are the discrepancies between the data and fitted distances. Assume that we have provisional values for all coordinates and that we wish to improve the typical coordinate, x_{kb}. If the problem of error is ignored, the solution of the following set of equations in the coordinates is suggested,

$$f(x_{kb}) = *d_{kj} - d_{kj} = *d_{kj} - \left[\sum_{a=1}^{m} (x_{ka} - x_{ja})^2 \right]^{1/2} = 0,$$

for $j \neq k = 1, 2, \ldots, n$. The use of Newton's method (e.g., Luenberger, 1984) for solving such equations leads to the following iterations (indexed by t), where the prime denotes the first derivative with respect to the kb-th coordinate:

$$\begin{aligned} x_{kb}^{t+1} &= x_{kb}^t - \frac{f(x_{kb}^t)}{f'(x_{kb}^t)} \\ &= x_{kb}^t + \frac{(*d_{kj} - d_{kj}^t)d_{kj}^t}{x_{kb}^t - x_{jb}^t} \\ &= x_{kb}^t + {}_jg_{kb}^t. \end{aligned}$$

For any coordinate, x_{kb}^t, there are several possible corrections, ${}_jg_{kb}^t$, one for each $j \neq k$. Since the data are typically not free of errors, some of these corrections will be too large and others too small. Therefore the corrections are averaged. The median yields an averaged correction that is not influenced by outliers in the data, in exactly the same way that the univariate median is robust. The median is taken over all $j \neq k$:

$$\begin{aligned} x_{kb}^{t+1} &= x_{kb}^t + \underset{j \neq k}{\text{med}} \, ({}_jg_{kb}^t) \\ &= x_{kb}^t + g_{kb}^t, \end{aligned}$$

where g_{kb}^t is used to denote the median of the possible corrections. As usual, some modification of the step size in the direction of the vector of corrections will speed convergence. The step size is computed as

$$\beta^t = \frac{\alpha^t}{g^t},$$

where g^t is the relative magnitude of the vector of corrections:

$$g^t = \left[\frac{\sum_{i,a} (g_{ia}^t)^2}{\sum_{i,a} (x_{ia}^t)^2} \right]^{1/2}.$$

The quantity α^t may be chosen in several ways. One choice that has been found to be reasonably satisfactory, and does not require any function evaluation, is based on a multi-variable generalization of Aitken's δ^2 method (Ramsay, 1975):

$$\alpha^{t+1} = \alpha^t \left[\frac{\sum_{i,a} (x_{ia}^{t-1} - x_{ia}^{t-2})^2}{\sum_{i,a} (x_{ia}^t - 2x_{ia}^{t-1} + x_{ia}^{t-2})^2} \right]^{1/2},$$

where α is revised after every third iteration. Hence the algorithm is given by

$$x_{kb}^{t+1} = x_{kb}^t + \beta^t g_{kb}^t.$$

The above is easily modified to accommodate changes in the model. For example, if an additive constant is allowed:

$$^*d_{ij} = d_{ij} - c.$$

This requires the following modification of the algorithm

$$x_{kb}^{t+1} = x_{kb}^t + \beta^t \operatorname*{med}_{j \neq k} \left[\frac{(^*d_{kj} - d_{kj}^t + c^t) d_{kj}^t}{(x_{kb}^t - x_{jb}^t)} \right],$$

and,

$$c^{t+1} = c^t - \operatorname*{med}_i \operatorname*{med}_{j \neq i} (^*d_{ij} - d_{ij}^t + c^t)$$

$$= \operatorname*{med}_i \operatorname*{med}_{j \neq 1} (^*d_{ij} - d_{ij}^t).$$

For the additive constant correction, either "$\text{med}_i \, \text{med}_{j \neq i}$" or "$\text{med}_{i<j}$" may be used. The former is somewhat more economical of storage. Note that no step size adjustment is used with the correction to the additive constant—extensive experimentation has shown this to be unnecessary.

4. A Starting Configuration

Although the algorithm described above generally performs reasonably well when started from a randomly selected position, there is no doubt that the likelihood of being trapped in a locally suboptimal position is much less if the starting configuration is close to the optimum. The suboptimal solution problem is ubiquitous with iterative methods and practical experience in many situations has shown that the best insurance against the problem is to have a starting position that is not too far from the optimum.

The classic Young-Householder-Torgerson scaling solution provides a poor start. As shown by the illustration in section 2, the procedure is very sensitive to the effect of outliers, and indeed turns out to be a poorer starting position than a random configuration. The starting configuration that is used here is simple and may be quickly computed. First, the data are replaced by their ranks, and then traditional Young-Householder-Torgerson scaling is performed. Subsequently, the scale of the recovered configuration is adjusted such that

$$\underset{i<j}{\text{med}}\ (*d_{ij} + c^0) = \underset{i<j}{\text{med}}\ (d_{ij}^0),$$

where the superscript zero indexes iteration zero, and c^0 is a robust starting estimate of the additive constant suggested by Torgerson's (1958) unidimensional subspace method:

$$c^0 = \underset{i}{\text{med}}\ \underset{j\neq i}{\text{med}}\ \underset{l\neq i,j}{\text{med}}\ (*d_{ij} - *d_{il} - *d_{jl}).$$

5. A Robust Index of Fit

Several measures of fit could be devised. The following bears an obvious and natural relationship to the iterative method, and has a simple interpretation, if multiplied by 100, as the median percentage discrepancy between the input data and the distances recovered. The index is called TUF to denote the idea of toughness or resistance in the face of badly behaved errors:

$$\text{TUF} = \underset{i}{\text{med}}\ \underset{j\neq i}{\text{med}} \left(\left| \frac{1 - d_{ij}}{*d_{ij}} \right| \right).$$

6. The Simulation

We examined the performance of several programs for multidimensional scaling in situations where the data were contaminated by the presence of outliers. This was done for two reasons. First, we wanted to see whether existing procedures would be adversely affected by outliers, and second, we wished to establish some basis for an evaluation of the present method. The comparisons described below are for the purposes of illustration only and are not intended to provide a comprehensive comparative evaluation of the different algorithms examined. The simulation is restricted in scope, and it is entirely possible that under other conditions the relative performance of the programs could be rather different. In addition, programs undergo changes over the years, and may include different features and rely on different algorithms by the time the reader sees this article.

The programs used were: (i) traditional Young-Householder-Torgerson metric scaling (Young & Householder, 1938; Torgerson, 1958), (ii) KYST-2 (Kruskal, Young, & Seery, 1978), in both metric (M) and nonmetric (NM) modes, (iii) ALSCAL-4 (Takane, Young, & de Leeuw, 1977; Young & Lewyckyj, 1981) in only its nonmetric model (NM), (iv) MULTISCALE II (Ramsay, 1977; 1982), and (v) our program TUFSCAL (for tough scaling). Some of these procedures are capable of fitting more than one model, and one program (MULTISCALE) even permits the specification of different error distributions. However, in order to make the simulation as similar as possible across different procedures, the data were generated using a linear model with no intercept. All procedures can fit this model, and indeed, classical metric scaling requires it. Notwithstanding, all procedures, except Young-Householder-Torgerson metric scaling, must estimate constants of scale, location, and even curvature, but this should not significantly affect their ability to recover configurations where the relationship between the data and the distances if of the form $*d_{ij} = d_{ij} + e_{ij}$.

It is not possible to choose the error distribution in a completely equitable fashion. Only MULTISCALE is explicit with regard to error: either lognormal or normal density functions on the distances may be selected. TUFSCAL, as a median based estimation procedure, requires no strong assumptions regarding the distribution of errors. The Young-Householder-Torgerson procedure should perform best when the errors are Gaussian on the scalar products, since the least squares fit is to the scalar products, but it is not clear what this implies about the error distribution on the distances. KYST(M) should perform well when the errors on the distances are Gaussian, since the distances are fit directly using least squares. However, in the case of nonmetric scaling routines such as KYST(NM) or ALSCAL(NM), it is difficult to know what conditions are to be preferred, with respect to the distribution of errors. Hence the choice of error model for the simulations was arbitrary: the lognormal was chosen as the error distribution because it is a plausible model for many psychological situations (Ramsay, 1977; 1982), and in many cases may be more appropriate than the Hefner model (Zinnes & MacKay, 1983) which has been very popular in simulation studies over the last 20 years.

The experiment was similar to many previous Monte Carlo simulations in the area of multidimensional scaling (Spence, 1983) and was comprised of two separate studies. In the first, only one known configuration was used in all conditions. This consisted of 21 points in two dimensions arranged in the fashion of a cross; the arms of the cross were each 1.0 units long, and the five individual points on any arm were equidistant from each neighboring point. The center point was at the origin of the space. The reason for choosing this arrangement was for ease of graphical presentation of the results.

The experimental design for the first study was a $6 \times 5 \times 2$ factorial layout with five replications per cell. The factors were:

A. Procedure: ALSCAL(NM), KYST(M), KYST(NM), MULTISCALE, TUFSCAL, or Y-H-T. All default stopping criteria were set to values that were unattainable, except in the case of perfect fit to the data, and thus the default maximum number of iterations permitted usually determined the stopping point.

B. Percentage of Outliers: 0, 5, 10, 20, or 40 percent of the 210 distances were contaminated.

C. Background Error: either zero or a moderate amount—see below. The data input to each of the procedures was generated as follows:

1. Euclidean interpoint distances among the points in the cross were calculated.

2. Distances were classified as outliers, or not, by randomly shuffling the distances (Knuth, 1969, p. 125, Algorithm P) and then selecting the first p percent to be outliers,

where p varied between 0 and 40, as noted above. All simulations used the random uniform generator GGUBFS (IMSL, 1980).

3. Pseudo random lognormal error was added to the distances by multiplying each true distance by an exponentially transformed pseudo random Gaussian generated by the polar method (Knuth, 1969, p. 104, Algorithm P; see also Ramsay, 1977). If a distance was classified as an outlier, the error standard deviation was equal to 2.0. For the remaining distances the error standard deviation was either zero (in the no background error condition), or 0.4 (in the moderate background error condition).

Each procedure received exactly the same input data and $6 \times 5 \times 2 \times 5 = 300$ separate scalings were performed, in two dimensions only.

The purpose of the second study was to check that the results of the first study were not idiosyncratic to the type of configuration used, and also to examine the behavior of TUFSCAL in other dimensionalities. The experimental design for the second study was similar to the first in many respects, and consisted of a $2 \times 3 \times 2 \times 3$ factorial with three replications:

A. Procedure: KYST(NM), TUFSCAL.

B. Percentage of Outliers: 0, 10, or 30 percent of the distances were contaminated.

C. Background Error: either zero or a moderate amount—as in the first study.

D. Configuration: Random configurations of either 10 points in one dimension, 20 points in two dimensions, or 30 points in three dimensions were used. The configurations were randomly generated by sampling the coordinates on each dimension from a rectangular distribution on the range $(-1, +1)$, subject to the constraint that no point be more than one unit from the origin. Different random configurations were used in each of the three replications.

The data input to both procedures was generated in exactly the same fashion as in the first study—see Points 1. through 3. above.

7. Results

Recovery was assessed by computing the Pearson correlation between the known distances and the distances recovered by each of the programs. Figures 2 and 3 display the mean recovery over five replications for each of the experimental conditions in the first study.

Figure 2 shows the performance of the procedures when there is no background error, but with a varying percentage of outliers. It can be seen that all programs recovered the known configuration when no outliers were present, but were less successful when the data were contaminated. Metric approaches were more adversely affected than nonmetric, with the worst performer being Young-Householder-Torgerson. This is essentially a least squares procedure that operates on *squared* dissimilarities, and hence the effect of any outlier is squared, with consequent high leverage on the solution. Metric procedures that operate directly on the dissimilarities, such as MULTISCALE and KYST(M) perform much better, but not as well as algorithms that employ a ranking transformation of the data, such as ALSCAL(NM) and KYST(NM). With up to 20 percent outliers, TUFSCAL does very well but its average performance falls off somewhat at 40 percent.

In Figure 3 the influence of background error is seen. As expected, all procedures perform more poorly, but, with one exception, the same general pattern obtains as with no background error. The exception is MULTISCALE, whose performance is comparable to KYST(NM) and ALSCAL(NM), at least up to 20 percent outliers.

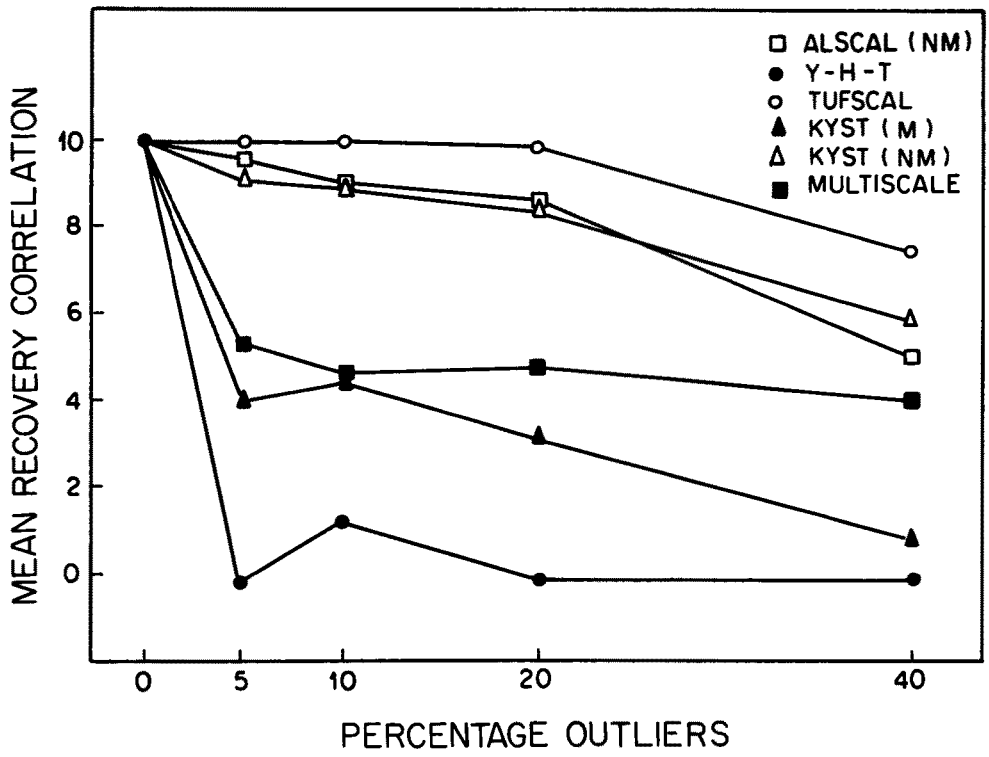

FIGURE 2
Recovery as a function of the percentage of outliers with no background error (correlation decimal point suppressed).

Averaging results over replications can hide what is possible with particular patterns of outliers. Consider, for example, what would happen if, by chance, *all* outliers happened to be in a single row/column of the data matrix. It is obvious that no procedure could recover the associated point with any accuracy. The corollary of this intuitive idea suggests that with certain patterns of outliers all procedures will perform better than with others. But a pattern of outliers that is not relatively homogeneous is likely to lead to poorer performance for any algorithm.

It is interesting to see what can happen under favorable conditions, and to this end we display the results of replication #2 in the no background error condition. Most programs did best in this replication, and this is presumably due to the fact that the common disposition of outliers was favorable in the sense discussed above. The recovered configurations are shown in Figure 4. It should be noted that the configurations presented are not necessarily comparable in size. For example, with Young-Householder-Torgerson in any outlier condition, the largest recovered distances are very much larger than the corresponding true distances, and the same problem is present, in lesser degree, with other programs also. If all configurations were drawn to scale, the overall display would be difficult to present neatly. Since we are really only interested in the locations of points relative to one another, all configurations have been scaled to fit within panels of the same size. This does result in some peculiar looking configurations where several points appear to have collapsed together. Zinnes and MacKay (1983) and Edwards (1986) have observed the same effect in a different context, but the cause is similar.

It is clear that even with 40 percent outliers present, the performance of TUFS-

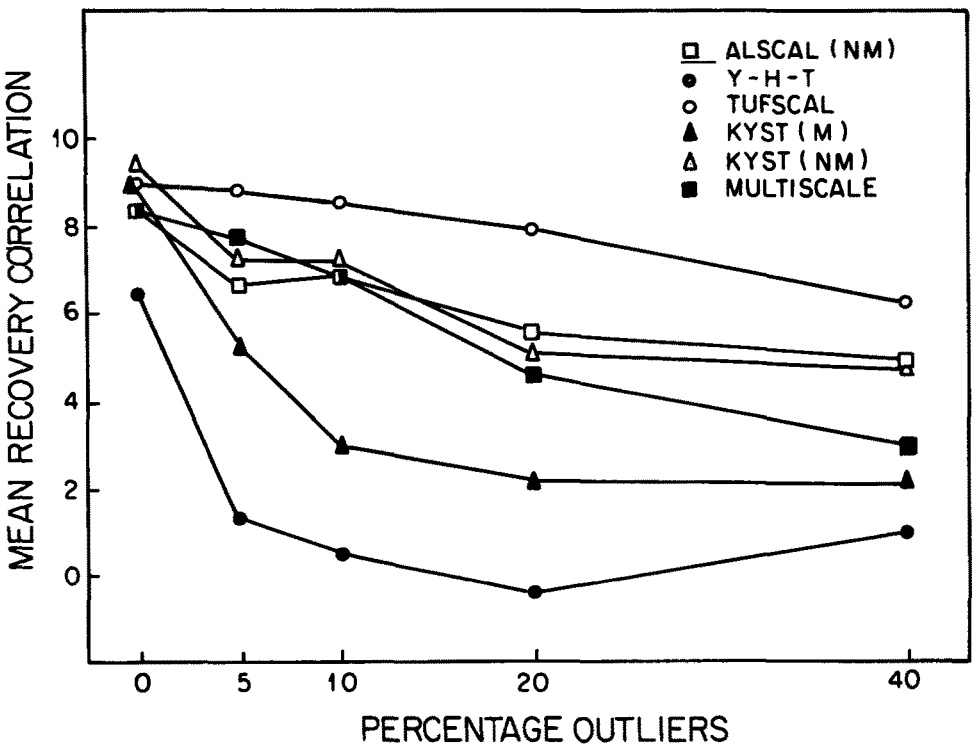

FIGURE 3
Recovery as a function of the percentage of outliers with moderate background error (correlation decimal point suppressed).

CAL is quite good. Although there are slight irregularities at the higher percentages, this does not necessarily reflect a basic inability to achieve perfect recovery when the percentage of outliers is high. One possible reason may be that a marginally suboptimal solution has been obtained, although it is probably more likely that, with the particular pattern of outliers in replication #2, perfect recovery is not attainable. We make this observation because, in other experiments, TUFSCAL has sometimes obtained a perfect reconstruction of the configuration when the percentage of outliers has exceeded 40 percent. Assigning a breakdown value to TUFSCAL is difficult, but it is probably safe to say that the breakdown value will exceed the percentage of outliers found in most data sets.

It is interesting to see how the actual recovered configurations shown in Figure 4 are related to the recovery correlations, and these are provided in Table 1. Configurations with associated recovery correlations of less than about 0.7 bear scant resemblance to the true configuration, and indeed the quality of recovery is not particularly impressive whenever the correlation fails to exceed 0.9.

It has been suggested (J. O. Ramsay, personal communication, August 18, 1985) that the performance of some of the programs might have been due, at least in part, to the fact that their starting configurations were unsatisfactory when outliers were present. All the programs employ some variant of the Young-Householder-Torgerson procedure, but only TUFSCAL replaces the data by their ranks. Since, as shown above, the unmodified classical metric scaling procedure does not cope well with data containing outliers, Ramsay's argument is plausible. The following demonstration shows that the quality of the start undoubtedly has some effect, but that this alone cannot account for the observed differences in performance.

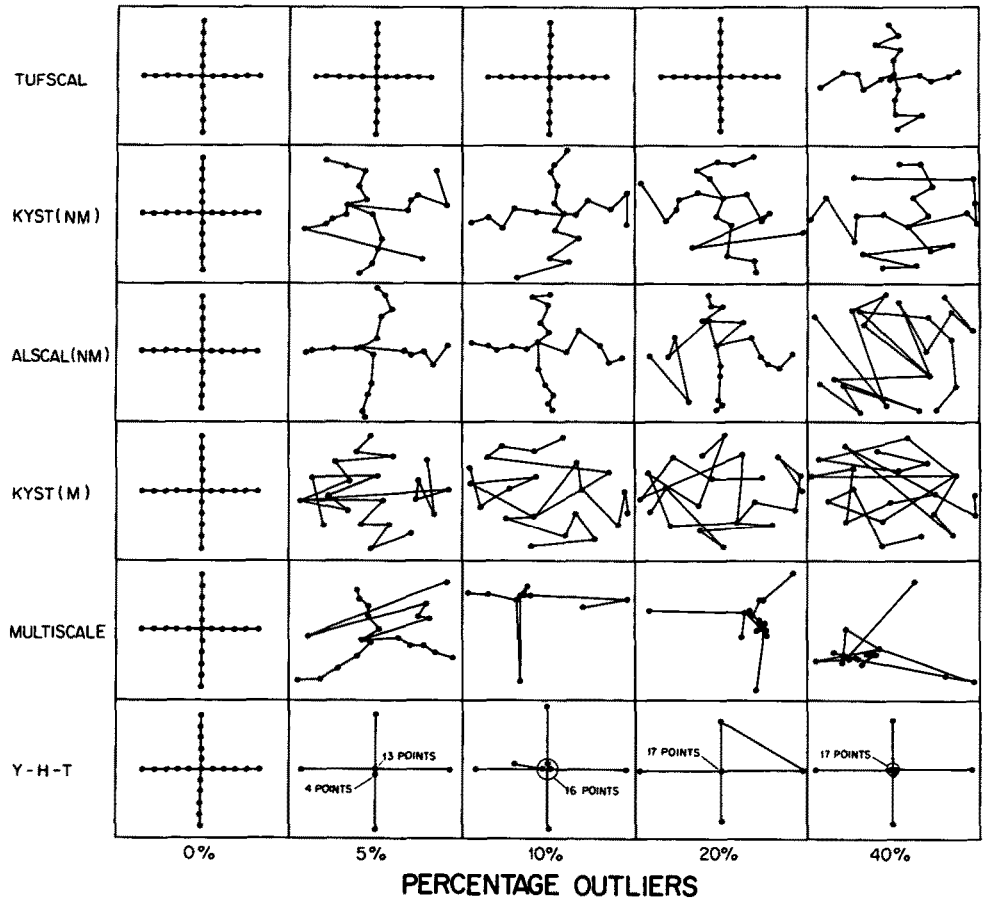

FIGURE 4
Recovered configurations in Replication #2.

Using KYST(NM) and TUFSCAL, we repeated replication Number 4 in the no background error condition. In this replication, TUFSCAL did well at all outlier percentage levels except the highest, where TUFSCAL and KYST did not differ much in terms of recovery correlation. We provided both programs with the best possible starting configuration—the true configuration—and the results are shown in Table 2. It can be seen that KYST's performance is indeed improved, on average, when a perfect start is provided, but with higher outlier percentages the recovery still falls short of perfect. This suggests that traditional algorithms will benefit from a more robust start when outliers are present. TUFSCAL did not move significantly from the perfect start, and achieved essentially perfect recovery in all outlier conditions. This shows that the original solution with 40 percent outliers was suboptimal, and suggests that there is room for improvement in both the TUFSCAL starting position and the iterative algorithm.

In the second study, solutions were obtained using TUFSCAL and KYST in one, two, and three dimensions, with random underlying configurations. The results are summarized in Table 3. The results are not at variance with the general pattern obtained in the first study, suggesting that the performance of programs is not idiosyncratic to a particular type of configuration, or particular number of dimensions. It should be noted that both procedures perform slightly worse as the number of coordinates estimated increases, in the presence of error. This is not surprising and is consistent with previous

Table 1

Recovery Correlations for Replication #2 in the No Background Error Condition

Procedure	Percentage Outliers				
	0	5	10	20	40
TUFSCAL	100	100	100	100	96
KYST(NM)	100	85	95	80	64
ALSCAL(NM)	100	96	95	81	49
MULTISCALE	100	87	39	53	20
KYST(M)	100	60	56	19	11
Y-H-T	100	02	-19	-21	06

Note:--Decimal points suppressed; all results rounded to 2 digits

Monte Carlo work (e.g., Young, 1970; Spence, 1972). However, when there are no outliers but error in the data, KYST performs slightly better than TUFSCAL with an increasing number of points and dimensions. The difference is small and, at the moment, we have no good explanation for it. In one dimension, with 0.4 error and 10 percent outliers, KYST appears to perform quite poorly (0.70 recovery correlation): this is mainly due to a probable local minimum solution among the three replications.

Table 2

Recovery Correlations in Replication #4 with Default and Perfect Starts in the No Background Error Condition

Procedure	Start	Percentage Outliers				
		0	5	10	20	40
TUFSCAL	Default	100	100	100	100	80
	Perfect	100	100	100	100	100
KYST(NM)	Default	100	97	98	64	73
	Perfect	100	97	94	89	93

Note:--Decimal points suppressed; all results rounded to 2 digits

Table 3

Mean Recovery Correlations with Random Configurations

Points/Dimensions	Procedure	Error S.D	Percentage Outliers		
			0	10	30
10/1	TUFSCAL	0.0	100	100	100
		0.4	97	96	94
	KYST(NM)	0.0	100	93	87
		0.4	97	70	81
20/2	TUFSCAL	0.0	100	100	98
		0.4	91	81	64
	KYST(NM)	0.0	100	86	71
		0.4	94	87	48
30/3	TUFSCAL	0.0	100	100	91
		0.4	80	75	53
	KYST(NM)	0.0	100	61	53
		0.4	86	56	38

Note:--Decimal points suppressed; all results rounded to 2 digits

8. Discussion

Of traditional programs, nonmetric methods seem to be most resistant to the influence of outliers. Metric procedures generally perform worse, although MULTISCALE is almost as resistant as a nonmetric procedure, when the simulated data contain a moderate amount of error. Young-Householder-Torgerson scaling and KYST(M) are quite badly affected and probably should not be used in situations where outliers are likely. At least in the conditions investigated in our experiments, the method implemented in TUFSCAL seems to be fairly resistant.

In view of the fact that traditional two way scaling programs may be adversely affected by relatively small numbers of aberrant observations, it may sometimes be helpful to use a resistant procedure for fitting the model. In situations where the data are not expected to be squeaky clean, the use of a robust method as an adjunct and complement to a traditional technique may be beneficial. Also, after a robust fit has been obtained, it is often rather easier to identify bad data points via residual analysis than when a conventional procedure has been used.

References

Edwards, D. R. (1986, June). Graphical artifacts in multidimensional scaling. Paper presented at a meeting of the Psychometric Society, Toronto.

Heiser, W. J. (1987, July). Multidimensional scaling with least absolute residuals. Presented at the First Conference of the International Federation of Classification Societies, Aachen, Federal Republic of Germany.

Huber, P. J. (1981). *Robust statistics*. New York: Wiley.

IMSL (1980). *IMSL library reference manual* (8th ed.). Houston, TX: Author.

Knuth, D. E. (1969). *The art of computer programming: Seminumerical algorithms* (Vol. 2). Reading, MA: Addison-Wesley.

Kruskal, J. B., Young, F. W., & Seery, J. B. (1978). *How to use KYST-2, a very flexible program to do multidimensional scaling and unfolding*. Murray Hill, NJ: Bell Laboratories.

Luenberger, D. G. (1984). *Linear and nonlinear programming* (2nd ed.). Reading, MA: Addison-Wesley.

Mosteller, F. & Tukey, J. W. (1977). *Data analysis and regression*. Reading, MA: Addison-Wesley.

Null, C. H., & Sarle, W. (1982, June). Robust multidimensional scaling. Paper presented at the Joint Meeting of the Psychometric and Classification Societies, Montreal, Canada.

Ramsay, J. O. (1975). Solving implicit equations in psychometric data analysis. *Psychometrika, 40*, 337–360.

Ramsay, J. O. (1977). Maximum likelihood estimation in multidimensional scaling. *Psychometrika, 42*, 241–266.

Ramsay, J. O. (1982). *Multiscale II manual* (Updated 1986). Montreal: McGill University.

Spence, I. (1972). A Monte Carlo evaluation of three nonmetric multidimensional scaling algorithms. *Psychometrika, 37*, 461–486.

Spence, I. (1982, June). Robust multidimensional scaling. Paper presented at the Joint Meeting of the Psychometric and Classification Societies, Montreal, Canada.

Spence, I. (1983). Monte Carlo simulation studies. *Applied Psychological Measurement, 7*, 405–425.

Takane, Y., Young, F. W., & de Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika, 42*, 7–67.

Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.

Young, F. W. (1970). Nonmetric multidimensional scaling: Recovery of metric information. *Psychometrika, 35*, 455–473.

Young, F. W., & Lewyckyj, R. (1981). *ALSCAL 4 user's guide*. Chapel Hill, NC: Data Analysis and Theory Associates.

Young, G., & Householder, A. S. (1938). Discussion of a set of points in terms of their mutual distances. *Psychometrika, 3*, 19–21.

Zinnes, J. L., & MacKay, D. B. (1983). Probabilistic multidimensional scaling: Complete and incomplete data. *Psychometrika, 48*, 27–48.

Manuscript received 10/2/83
Final version received 4/12/88